

A Simple Method for Obtaining the Maximal Correlation Coefficient and Related Characterizations*

Nickos Papadatos[†] and Tatiana Xifara[‡]

Section of Statistics and O.R., Department of Mathematics, University of Athens,
Panepistemiopolis, 157 84 Athens, Greece

and

Department of Mathematics and Statistics, Lancaster University, UK

Abstract

We provide a method that enables the simple calculation of the maximal correlation coefficient of a bivariate distribution, under suitable conditions. In particular, the method readily applies to known results on order statistics and records. As an application we provide a new characterization of the exponential distribution: Under a splitting model on iid observations, it is the (unique, up to a location transformation) parent distribution that maximizes the correlation coefficient between the records among two different branches of the splitting sequence.

MSC: Primary 60E15.

Key words and phrases: Maximal Correlation Coefficient; Characterization of Exponential Distribution; Order Statistics; Records; Splitting model.

1 Introduction

As is well-known, the *Pearson correlation coefficient* of the random variables X and Y is defined as

$$\rho(X, Y) = \text{Corr}(X, Y) = \frac{\text{Cov}(X, Y)}{\sqrt{\text{Var}(X)}\sqrt{\text{Var}(Y)}},$$

provided that $0 < \text{Var}(X) < \infty$ and $0 < \text{Var}(Y) < \infty$. It assumes values in the interval $[-1, 1]$ and it is a measure of *linear dependence* of X and Y . Although

*Work partially supported by the University of Athens Research Grant 70/4/5637.

[†]*Corresponding author.* e-mail: npapadat@math.uoa.gr url: <http://users.uoa.gr/~npapadat/>

[‡]e-mail: t.xifara@lancaster.ac.uk url: <http://www.maths.lancs.ac.uk/~xifara/>

$\rho(X, Y) = 0$ for independent X and Y , the converse is not true. Gebelein (1941) introduced the *maximal correlation coefficient*,

$$R(X, Y) = \sup \text{Corr}(g_1(X), g_2(Y)),$$

where the supremum is taken over all Borel functions $g_1 : \mathbb{R} \rightarrow \mathbb{R}$ and $g_2 : \mathbb{R} \rightarrow \mathbb{R}$ with $0 < \text{Var} g_1(X) < \infty$ and $0 < \text{Var} g_2(Y) < \infty$. In contrast to $\rho(X, Y)$, $R(X, Y)$ is defined whenever both X and Y are non-degenerate, assumes values in the interval $[0, 1]$ and vanishes if and only if X and Y are independent. The maximal correlation coefficient plays a fundamental role in various areas of statistics; e.g., it is useful in obtaining optimal transformations for regression, Breiman and Friedman (1985), and it has applications in the convergence theory of Gibbs sampling algorithms, Liu et al. (1994).

However, despite its usefulness, it is often difficult to calculate the maximal correlation coefficient in an explicit form, except in some rare cases. A well-known exception is the result of Gebelein (1941) and Lancaster (1957) who showed the interesting property that if (X, Y) is bivariate normal then

$$R(X, Y) = |\text{Corr}(X, Y)|. \quad (1)$$

Another exception is provided by the surprising result of Dembo et al. (2001), and its subsequent extensions given by Bryc et al. (2005) and Yu (2008). In its general form the result states that for any iid non-degenerate random variables X_1, \dots, X_n ,

$$R(X_1 + \dots + X_m, X_{k+1} + \dots + X_n) = \frac{m - k}{\sqrt{m(n - k)}}, \quad 1 \leq k + 1 \leq m \leq n.$$

Finally, we mention an important result of Székely and Móri (1985), who showed, using Jacobi polynomials, that if (X, Y) follows a bivariate density of the form

$$f(x, y) = \frac{\Gamma(\alpha + \beta + \gamma)}{\Gamma(\alpha)\Gamma(\beta)\Gamma(\gamma)} x^{\alpha-1} (y - x)^{\beta-1} (1 - y)^{\gamma-1}, \quad 0 < x < y < 1 \quad (2)$$

(where the parameters α, β, γ are positive), then

$$R(X, Y) = \text{Corr}(X, Y) = \sqrt{\frac{\alpha\gamma}{(\beta + \alpha)(\beta + \gamma)}}. \quad (3)$$

Observe that for any integers $1 \leq i < j \leq n$, the density of the pair of order statistics $(U_{i:n}, U_{j:n})$, based on n iid rv's observations from the standard uniform distribution, is of the form (2) (with $\alpha = i, \beta = j - i, \gamma = n + 1 - j$). Actually, (3) extends Terrell's (1983) characterization of rectangular distributions through maximal correlation of an ordered pair.

In this article we provide a unified method for obtaining the maximal correlation when the bivariate distribution has a particular structure (diagonal structure – see next section). The method is very simple (e.g., it readily applies to verify (1) and

(3)) and it does not require knowledge of particular sets of orthogonal polynomials. As notable examples, some known related characterizations of specific distributions through maximal correlation of ordered data and records are presented in Section 3. Finally, in Section 4 we consider a splitting model based on iid observations. Applying this method it is shown that the records among two different branches of the splitting sequence are maximally correlated if and only if the population distribution is exponential (up to a location transformation) – this fact extends a characterization of Nevzorov (1992).

2 The maximal correlation coefficient of bivariate distributions having diagonal structure

Let (X, Y) be an arbitrary random vector with distribution function $F(x, y)$ and assume that both X and Y are non-degenerate. We say that F (or the vector (X, Y)) has diagonal structure if the following three conditions are satisfied.

A1. We assume that both X and Y have all their moments finite:

$$\mathbb{E}|X|^n < \infty \text{ and } \mathbb{E}|Y|^n < \infty \text{ for } n = 1, 2, \dots \quad (4)$$

It is known that, under (4), there exists a (unique) orthonormal polynomial system (OPS) $\{\phi_n(x) = p_n x^n + \text{Pol}_{n-1}(x), p_n > 0, n = 0, 1, \dots\}$, corresponding to X , and a (unique) OPS $\{\psi_n(y) = q_n y^n + \text{Pol}_{n-1}(y), q_n > 0, n = 0, 1, \dots\}$, corresponding to Y , where $\phi_0(x) \equiv \psi_0(y) \equiv 1$ and $\text{Pol}_k(t)$ denotes an arbitrary polynomial in t of degree less than or equal to k , that may changes from line to line. The orthonormality of the above OPS's means, as usual, that

$$\mathbb{E}[\phi_n(X)\phi_k(X)] = \mathbb{E}[\psi_n(Y)\psi_k(Y)] = \delta_{nk}, \quad k, n = 0, 1, \dots,$$

where δ_{nk} is Kronecker's δ . It should be noted that the OPS for X reduces to a finite set, say $\{\phi_n(x)\}_{n=0}^N$, if and only if the support of X is concentrated on a finite subset of \mathbb{R} having $N + 1 \geq 2$ points; the same is true for the OPS of Y .

A2. We assume that the OPS $\{\phi_n(x)\}_{n=0}^\infty$ is complete in $L^2(X)$, the Hilbert space of all Borel functions $g : \mathbb{R} \rightarrow \mathbb{R}$ with $\text{Var}g(X) < \infty$ (note that two functions g_1, g_2 are considered as “equal” if $\mathbb{P}[g_1(X) = g_2(X)] = 1$). Similarly, we assume that the system $\{\psi_n(y)\}_{n=0}^\infty$ is complete in $L^2(Y)$.

A3. We assume that the random vector (X, Y) has the *polynomial regression property*, that is,

$$\begin{aligned} \mathbb{E}(X^n|Y) &= A_n Y^n + \text{Pol}_{n-1}(Y), \quad n = 1, 2, \dots, \\ \mathbb{E}(Y^n|X) &= B_n X^n + \text{Pol}_{n-1}(X), \quad n = 1, 2, \dots, \end{aligned}$$

where $A_n, B_n \in \mathbb{R}$.

The assumptions A1 and A2 are not restrictive since, e.g., they are satisfied whenever both X and Y have finite moment generating functions in a neighborhood

of 0; see, e.g., Koudou (1998) and Afendras et al. (2011). However, this is not the case for assumption A3, since it applies to very particular distributions, as the following lemma shows.

Lemma 2.1. Using the above notation and assuming A1–A3 we have that for all $n, k \in \{1, 2, \dots\}$,

$$\mathbb{E}[\phi_n(X)\psi_k(Y)] = \delta_{nk}\rho_n, \quad (5)$$

where δ_{nk} is Kronecker's delta and $\rho_n = \mathbb{E}[\phi_n(X)\psi_n(Y)] \in [-1, 1]$.

Proof: Since $\phi_n(X)$ and $\psi_n(Y)$ are standardized random variables, we have $\rho_n = \text{Corr}[\phi_n(X), \psi_n(Y)]$ and, therefore, $\rho_n \in [-1, 1]$. Now, if $1 \leq k < n$ then A3 yields

$$\mathbb{E}[\phi_n(X)\psi_k(Y)] = \mathbb{E}\{\phi_n(X)\mathbb{E}[\psi_k(Y)|X]\} = \mathbb{E}[\phi_n(X)\text{Pol}_k(X)] = 0,$$

because ϕ_n is orthogonal to any polynomial of degree at most $n - 1$. Similar arguments apply to the case $1 \leq n < k$, and the proof is complete. \square

The bivariate distributions satisfying (5) are sometimes called *Lancaster distributions* and the correlations ρ_n form a *Lancaster sequence* with respect to X and Y ; see Lancaster (1969); cf., e.g., Koudou (1996, 1998). Therefore, by Lemma 2.1 we see that assumption A3 forces a distribution to be a Lancaster one. Under certain conditions, the density of a Lancaster distribution (if exists) has the formal representation (diagonal structure)

$$f(x, y) = f_X(x)f_Y(y) \left(1 + \sum_{n=1}^{\infty} \rho_n \phi_n(x)\psi_n(y) \right),$$

where f_X and f_Y are the marginal densities of X and Y .

Another useful observation is the following: If the assumptions A1–A3 are satisfied then we can calculate each ρ_n , and this calculation does not require any knowledge of the polynomial systems $\{\phi_n(x)\}_{n=0}^{\infty}$ and $\{\psi_n(y)\}_{n=0}^{\infty}$. Indeed, we have the following

Lemma 2.2. Using the above notation and assuming A1–A3 we have that for all $n \in \{1, 2, \dots\}$,

$$A_n B_n \geq 0, \quad \rho_n = \text{sign}(A_n) \sqrt{A_n B_n} \quad \text{and} \quad |\rho_n| = \sqrt{A_n B_n}. \quad (6)$$

Proof: Since $\phi_n(X) = p_n X^n + \text{Pol}_{n-1}(X)$ and $\psi_n(Y) = q_n Y^n + \text{Pol}_{n-1}(Y)$ we have

$$\begin{aligned} \rho_n &= \mathbb{E}\{\psi_n(Y)\mathbb{E}(\phi_n(X)|Y)\} = \mathbb{E}\{\psi_n(Y)[p_n \mathbb{E}(X^n|Y) + \text{Pol}_{n-1}(Y)]\} \\ &= p_n \mathbb{E}[\psi_n(Y)\mathbb{E}(X^n|Y)] + 0 = p_n \mathbb{E}\{\psi_n(Y)[A_n Y^n + \text{Pol}_{n-1}(Y)]\} \\ &= p_n A_n \mathbb{E}[\psi_n(Y)Y^n] + 0 = p_n A_n \mathbb{E}\{\psi_n(Y)q_n^{-1}[\psi_n(Y) - \text{Pol}_{n-1}(Y)]\} \\ &= \frac{p_n A_n}{q_n} \mathbb{E}[\psi_n^2(Y)] - 0 = \frac{p_n A_n}{q_n}. \end{aligned}$$

This shows that ρ_n and A_n have the same sign. Using the same arguments (conditioning on X) it follows that $\rho_n = \frac{q_n B_n}{p_n}$; thus, $\rho_n^2 = A_n B_n$, and the proof is complete. \square

We are now in a position to state and prove our main result.

Theorem 2.1. If the assumptions A1–A3 are satisfied then

$$R(X, Y) = \sup_{n \geq 1} |\rho_n| = \sup_{n \geq 1} \sqrt{A_n B_n}. \quad (7)$$

Moreover, if $|\rho_n| < |\rho_{n_0}|$ for all $n \geq 1$, $n \neq n_0$, then for any $g_1 \in L^2(X)$ with $\text{Var } g_1(X) > 0$ and for any $g_2 \in L^2(Y)$ with $\text{Var } g_2(Y) > 0$ we have the inequality

$$\text{Corr}[g_1(X), g_2(Y)] \leq |\rho_{n_0}|,$$

with equality if and only if $g_1(x) = a_0 + a_1 \phi_{n_0}(x)$ and $g_2(y) = b_0 + b_1 \psi_{n_0}(y)$ for some constants $a_0, b_0, a_1, b_1 \in \mathbb{R}$ with $a_1 b_1 \text{sign}(A_{n_0}) > 0$.

Proof: Let $g_1 \in L^2(X)$. By the completeness of $\{\phi_n\}_{n=0}^\infty$ it follows that g_1 admits the representation

$$g_1(x) = \sum_{n=0}^{\infty} \alpha_n \phi_n(x), \quad \text{where } \alpha_n = \mathbb{E}[g_1(X) \phi_n(X)] = \int_{\mathbb{R}} g_1(x) \phi_n(x) dF_X(x).$$

Here F_X is the marginal distribution of X , $\{\alpha_n\}_{n=0}^\infty$ are the Fourier coefficients with respect to the OPS $\{\phi_n\}_{n=0}^\infty$, and the series converges in the $L^2(X)$ -sense, i.e.,

$$\lim_N \mathbb{E} \left[g_1(X) - \sum_{n=0}^N \alpha_n \phi_n(X) \right]^2 = 0. \quad (8)$$

In particular, $\alpha_0 = \mathbb{E}[g_1(X)]$, and the above limit is usually written as Parseval's identity,

$$\text{Var } g_1(X) = \sum_{n=1}^{\infty} \alpha_n^2,$$

because it is easily seen that

$$\mathbb{E} \left[g_1(X) - \sum_{n=0}^N \alpha_n \phi_n(X) \right]^2 = \text{Var } g_1(X) - \sum_{n=1}^N \alpha_n^2.$$

Therefore, the assumption $\text{Var } g_1(X) > 0$ implies that $\alpha_n \neq 0$ for at least one $n \geq 1$. Similarly, for any $g_2 \in L^2(Y)$ we have

$$\text{Var } g_2(Y) = \sum_{n=1}^{\infty} \beta_n^2, \quad \text{where } \beta_n = \mathbb{E}[g_2(Y) \psi_n(Y)] = \int_{\mathbb{R}} g_2(y) \psi_n(y) dF_Y(y).$$

Here F_Y is the marginal distribution of Y , $\{\beta_n\}_{n=0}^\infty$ are the Fourier coefficients with respect to the OPS $\{\psi_n\}_{n=0}^\infty$ and, as before,

$$\lim_N \mathbb{E} \left[g_2(Y) - \sum_{n=0}^N \beta_n \psi_n(Y) \right]^2 = \text{Var } g_2(Y) - \lim_N \sum_{n=1}^N \beta_n^2 = 0. \quad (9)$$

Using the above we can show that

$$\mathbb{E}[g_1(X)\psi_n(Y)] = \alpha_n\rho_n \quad \text{and} \quad \mathbb{E}[g_2(Y)\phi_n(X)] = \beta_n\rho_n, \quad n = 1, 2, \dots \quad (10)$$

Indeed, for any $N \geq n$ we have

$$\mathbb{E}[g_1(X)\psi_n(Y)] = \mathbb{E}\left\{\left[g_1(X) - \sum_{k=0}^N \alpha_k \phi_k(X)\right] \psi_n(Y)\right\} + \sum_{k=0}^N \alpha_k \mathbb{E}[\phi_k(X)\psi_n(Y)].$$

Now since $N \geq n$, $\phi_0(x) \equiv 1$, $\mathbb{E}[\psi_n(Y)] = 0$, $\mathbb{E}[\psi_n^2(Y)] = 1$ and $\mathbb{E}[\phi_k(X)\psi_n(Y)] = \delta_{kn}\rho_n$ for $k \geq 1$, we conclude, in view of (8) and by using the Cauchy-Schwarz inequality, that

$$\begin{aligned} 0 &\leq (\mathbb{E}[g_1(X)\psi_n(Y)] - \alpha_n\rho_n)^2 = \left(\mathbb{E}\left\{\left[g_1(X) - \sum_{k=0}^N \alpha_k \phi_k(X)\right] \psi_n(Y)\right\}\right)^2 \\ &\leq \mathbb{E}\left[g_1(X) - \sum_{k=0}^N \alpha_k \phi_k(X)\right]^2 \mathbb{E}[\psi_n^2(Y)] \rightarrow 0, \quad \text{as } N \rightarrow \infty; \end{aligned}$$

therefore, since $(\mathbb{E}[g_1(X)\psi_n(Y)] - \alpha_n\rho_n)^2$ does not depend on N , we conclude the first identity in (10), while the second one follows by the same arguments. Using (10) it is easily seen that

$$\begin{aligned} &\mathbb{E}\left[\left(g_1(X) - \sum_{n=0}^N \alpha_n \phi_n(X)\right) \left(g_2(Y) - \sum_{n=0}^N \beta_n \psi_n(Y)\right)\right] \\ &= \text{Cov}[g_1(X), g_2(Y)] - \sum_{n=1}^N \rho_n \alpha_n \beta_n; \end{aligned}$$

thus, squaring the above identity and applying the Cauchy-Schwarz inequality in the resulting squared expectation we conclude, in view of (8) and (9), that

$$\text{Cov}[g_1(X), g_2(Y)] = \sum_{n=1}^{\infty} \rho_n \alpha_n \beta_n. \quad (11)$$

Therefore, combining the above we get the expression

$$\text{Corr}[g_1(X), g_2(Y)] = \frac{\sum_{n=1}^{\infty} \rho_n \alpha_n \beta_n}{\sqrt{\sum_{n=1}^{\infty} \alpha_n^2} \sqrt{\sum_{n=1}^{\infty} \beta_n^2}}. \quad (12)$$

Now observe that, in view of (11),

$$\begin{aligned}
(\text{Cov}[g_1(X), g_2(Y)])^2 &= \left| \sum_{n=1}^{\infty} \rho_n \alpha_n \beta_n \right|^2 \leq \left(\sum_{n=1}^{\infty} |\rho_n| |\alpha_n| |\beta_n| \right)^2 \\
&= \left(\sum_{n=1}^{\infty} (\sqrt{|\rho_n|} |\alpha_n|) (\sqrt{|\rho_n|} |\beta_n|) \right)^2 \\
&\leq \left(\sum_{n=1}^{\infty} |\rho_n| \alpha_n^2 \right) \left(\sum_{n=1}^{\infty} |\rho_n| \beta_n^2 \right) \\
&\leq \left(\left(\sup_{n \geq 1} |\rho_n| \right) \sum_{n=1}^{\infty} \alpha_n^2 \right) \left(\left(\sup_{n \geq 1} |\rho_n| \right) \sum_{n=1}^{\infty} \beta_n^2 \right) \\
&= \left(\sup_{n \geq 1} \rho_n^2 \right) \left(\sum_{n=1}^{\infty} \alpha_n^2 \right) \left(\sum_{n=1}^{\infty} \beta_n^2 \right).
\end{aligned}$$

The above inequality, combined with (12), shows that

$$R(X, Y) \leq \sup_{n \geq 1} |\rho_n| = R, \text{ say.}$$

On the other hand, for any $\epsilon > 0$ we can find an index n_0 such that $|\rho_{n_0}| > R - \epsilon$, and thus, $|\text{Corr}[\phi_{n_0}(X), \psi_{n_0}(Y)]| = |\rho_{n_0}| > R - \epsilon$. Therefore,

$$\begin{aligned}
R(X, Y) &= \sup \text{Corr}[g_1(X), g_2(Y)] \\
&\geq \max\{\text{Corr}[\phi_{n_0}(X), \psi_{n_0}(Y)], \text{Corr}[-\phi_{n_0}(X), \psi_{n_0}(Y)]\} \\
&= \max\{\rho_{n_0}, -\rho_{n_0}\} = |\rho_{n_0}| > R - \epsilon.
\end{aligned}$$

Since $\epsilon > 0$ is arbitrary it follows that $R(X, Y) \geq R$, and thus, $R(X, Y) = R$. Finally, it is obvious that if the sequence $\{|\rho_n|\}_{n=1}^{\infty}$ has a unique maximum, say $|\rho_{n_0}|$, then, working as above, it is easily seen that

$$(\text{Cov}[g_1(X), g_2(Y)])^2 \leq \rho_{n_0}^2 \left(\sum_{n=1}^{\infty} \alpha_n^2 \right) \left(\sum_{n=1}^{\infty} \beta_n^2 \right) = \rho_{n_0}^2 \text{Var } g_1(X) \text{Var } g_2(Y),$$

with equality if and only if $\alpha_n = \beta_n = 0$ for all $n \geq 1, n \neq n_0$; this, combined with the fact that $\rho_{n_0} (= \text{Corr}[\phi_{n_0}(X), \psi_{n_0}(Y)])$ has the sign of A_{n_0} , completes the proof. \square

3 Examples providing known characterizations via maximal correlation

The following known results are immediate applications of Theorem 2.1.

The bivariate normal case. Assumptions A1–A3 are easily checked for the bivariate normal. Indeed, if (X, Y) is bivariate normal with $\mathbb{E}(X) = \mu_1, \mathbb{E}(Y) = \mu_2, \text{Var}(X) =$

$\sigma_1^2 > 0$, $\text{Var}(Y) = \sigma_2^2 > 0$ and $\text{Corr}(X, Y) = \rho \in [-1, 1]$ then it is well-known that $(X|Y = y) \sim N(\mu_1 + \rho \frac{\sigma_1}{\sigma_2}(y - \mu_2), (1 - \rho^2)\sigma_1^2)$; this means that

$$(X|Y = y) \stackrel{d}{=} \mu_1 + \rho \frac{\sigma_1}{\sigma_2}(y - \mu_2) + \sigma_1 \sqrt{1 - \rho^2} Z,$$

where $Z \sim N(0, 1)$ and $\stackrel{d}{=}$ denotes equality in distribution. Therefore,

$$\mathbb{E}[X^n|Y = y] = \mathbb{E}[\mu_1 + \rho \frac{\sigma_1}{\sigma_2}(y - \mu_2) + \sigma_1 \sqrt{1 - \rho^2} Z]^n = \rho^n \frac{\sigma_1^n}{\sigma_2^n} y^n + \text{Pol}_{n-1}(y),$$

that is,

$$\mathbb{E}[X^n|Y] = A_n Y^n + \text{Pol}_{n-1}(Y), \text{ where } A_n = \rho^n \frac{\sigma_1^n}{\sigma_2^n}, \quad n = 1, 2, \dots$$

Similarly,

$$\mathbb{E}[Y^n|X] = B_n X^n + \text{Pol}_{n-1}(X), \text{ where } B_n = \rho^n \frac{\sigma_2^n}{\sigma_1^n}, \quad n = 1, 2, \dots$$

Thus, A3 is satisfied, while A1 and A2 are obvious. It follows from (6) that $|\rho_n| = \sqrt{A_n B_n} = |\rho|^n$, $\rho_n = \text{sign}(\rho^n) |\rho|^n = \rho^n$, and, by (7), $R(X, Y) = \sup_{n \geq 1} |\rho_n| = \max_{n \geq 1} |\rho|^n = |\rho|$; moreover, if $0 < |\rho| < 1$, the equality in the inequality

$$|\text{Corr}[g_1(X), g_2(Y)]| \leq |\rho|$$

holds if and only if both g_1 and g_2 are linear. On the other hand it is worth to note that (11) takes here the form (cf. Afendras et al. (2011))

$$\text{Cov}[g_1(X), g_2(Y)] = \sum_{n=1}^{\infty} \frac{\rho^n \sigma_1^n \sigma_2^n}{n!} \mathbb{E}[g_1^{(n)}(X)] \mathbb{E}[g_2^{(n)}(Y)], \quad (13)$$

provided that $g_1, g_2 \in C^\infty$, $g_1(X) \in L^2(X)$, $g_2(Y) \in L^2(Y)$, and that $\mathbb{E}[g_1^{(n)}(X)] < \infty$ and $\mathbb{E}[g_2^{(n)}(Y)] < \infty$ for all n , where $g_i^{(n)}$ denotes the n -th derivative of g_i , $i = 1, 2$. Of course, one can apply (13) to the case $X = Y$; then $\mu_1 = \mu_2 = \mu$, say, $\rho = 1$, $\sigma_1 = \sigma_2 = \sigma$, say, and (13) yields the generalized Stein identity for the $N(\mu, \sigma^2)$ distribution:

$$\text{Cov}[g_1(X), g_2(X)] = \sum_{n=1}^{\infty} \frac{(\sigma^2)^n}{n!} \mathbb{E}[g_1^{(n)}(X)] \mathbb{E}[g_2^{(n)}(X)].$$

Characterization of rectangular distributions via maximal correlation of order statistics. Terrell (1983), using Legendre polynomials, proved that if $X_{1:2} \leq X_{2:2}$ are the order statistics from two iid observations from a distribution with finite variance then

$$\text{Corr}(X_{1:2}, X_{2:2}) \leq \frac{1}{2},$$

and the equality characterizes the rectangular (uniform over some non-degenerate finite interval) distributions. However, Theorem 2.1 applies immediately here. Indeed, if $\mathcal{U}(a, b)$ denotes the uniform distribution over (a, b) and if $U_1, U_2 \sim \mathcal{U}(0, 1)$ then it is obvious that the order statistics $U_{1:2} \leq U_{2:2}$ satisfy the following:

$$\begin{aligned} U_{1:2}|U_{2:2} \sim \mathcal{U}(0, U_{2:2}) &\Rightarrow \mathbb{E}[U_{1:2}^n|U_{2:2}] = \int_0^{U_{2:2}} t^n \frac{1}{U_{2:2}} dt = \frac{1}{n+1} U_{2:2}^n, \\ U_{2:2}|U_{1:2} \sim \mathcal{U}(U_{1:2}, 1) &\Rightarrow \mathbb{E}[U_{2:2}^n|U_{1:2}] = \int_{U_{1:2}}^1 t^n \frac{1}{1-U_{1:2}} dt \\ &= \frac{1}{n+1} (1 + U_{1:2} + \cdots + U_{1:2}^n). \end{aligned}$$

Thus, $A_n = B_n = \frac{1}{n+1}$ and $|\rho_n| = \frac{1}{n+1}$. Therefore, $\max_{n \geq 1} |\rho_n| = |\rho_1| = \frac{1}{2}$. It follows from Theorem 2.1 that $\text{Corr}[g(U_{1:2}), g(U_{2:2})] \leq \frac{1}{2}$, with equality if and only if g is linear. Since for the order statistics $X_{1:2} \leq X_{2:2}$ from an arbitrary distribution F it is true that

$$(X_{1:2}, X_{2:2}) \stackrel{d}{=} (g(U_{1:2}), g(U_{2:2})), \text{ where } g(u) = \inf\{x : F(x) \geq u\}, 0 < u < 1,$$

(the above g is usually denoted as F^{-1}), Terrell's result follows. The above argument can be easily extended to provide the characterization of Székely and Móri (1985), who showed, using Jacobi polynomials, that for any integers $1 \leq i < j \leq n$,

$$\text{Corr}(X_{i:n}, X_{j:n}) \leq \sqrt{\frac{i(n+1-j)}{j(n+1-i)}},$$

with equality if and only if the random sample arises from a rectangular distribution. Indeed, setting $g(u) = F^{-1}(u) = \inf\{x : F(x) \geq u\}$, $0 < u < 1$, where F is the common distribution function of the iid rv's X_1, \dots, X_n , we have

$$(X_{i:n}, X_{j:n}) \stackrel{d}{=} (g(U_{i:n}), g(U_{j:n})) \text{ and, thus, } \text{Corr}(X_{i:n}, X_{j:n}) = \text{Corr}(g(U_{i:n}), g(U_{j:n})),$$

which is well defined whenever $0 < \text{Var } X_{i:n} + \text{Var } X_{j:n} < \infty$. Since for any $s \in (0, 1)$, $(U_{i:n}|U_{j:n} = s) \stackrel{d}{=} \tilde{U}_{i:j-1}$, where $\tilde{U}_{i:m}$ is the i -th order statistic from a sample of size m from $\mathcal{U}(0, s)$, we have

$$\tilde{U}_{i:j-1} \stackrel{d}{=} sU_{i:j-1} \Rightarrow \mathbb{E}[U_{i:n}^k|U_{j:n} = s] = \mathbb{E}[(sU_{i:j-1})^k] = s^k \mathbb{E}[U_{i:j-1}^k].$$

Now calculate

$$\begin{aligned} \mathbb{E}[U_{i:j-1}^k] &= \int_0^1 u^k \frac{1}{B(i, j-i)} u^{i-1} (1-u)^{j-i-1} du \\ &= \frac{B(k+i, j-i)}{B(i, j-i)} = \frac{(k+i-1)!(j-1)!}{(k+j-1)!(i-1)!}. \end{aligned}$$

Also, for any $t \in (0, 1)$ we have $(U_{j:n}|U_{i:n} = t) \stackrel{d}{=} \tilde{U}_{j-i:n-i}$, where $\tilde{U}_{j-i:n-i}$ is the $(j-i)$ -th order statistic from a sample of size $n-i$ from $\mathcal{U}(t, 1)$. Clearly, if $\tilde{U} \sim \mathcal{U}(t, 1)$ then

$\tilde{U} \stackrel{d}{=} t + (1-t)U$ where $U \sim \mathcal{U}(0, 1)$. Therefore, $(U_{j:n}|U_{i:n} = t) \stackrel{d}{=} t + (1-t)U_{j-i:n-i}$ and since $U_{j-i:n-i} \stackrel{d}{=} 1 - U_{n+1-j:n-i}$, we get $(U_{j:n}|U_{i:n} = t) \stackrel{d}{=} 1 - U_{n+1-j:n-i} + tU_{n+1-j:n-i}$. Therefore,

$$\begin{aligned} \mathbb{E}[U_{j:n}^k | U_{i:n} = t] &= \mathbb{E}[1 - U_{n+1-j:n-i} + tU_{n+1-j:n-i}]^k \\ &= t^k \mathbb{E}[U_{n+1-j:n-i}^k] + \text{Pol}_{k-1}(t) \\ &= t^k \int_0^1 u^k \frac{1}{B(n+1-j, j-i)} u^{n-j} (1-u)^{j-i-1} du + \text{Pol}_{k-1}(t) \\ &= \frac{(n+k-j)!(n-i)!}{(n+k-i)!(n-j)!} t^k + \text{Pol}_{k-1}(t). \end{aligned}$$

Thus, we found that assumption A3 is satisfied with

$$A_k = \frac{(k+i-1)!(j-1)!}{(k+j-1)!(i-1)!} = \frac{[i]_k}{[j]_k},$$

where $[\alpha]_k = \alpha(\alpha+1) \cdots (\alpha+k-1)$, and

$$B_k = \frac{(n+k-j)!(n-i)!}{(n+k-i)!(n-j)!} = \frac{[n+1-j]_k}{[n+1-i]_k}.$$

Hence,

$$\rho_k^2 = A_k B_k = \frac{[i]_k [n+1-j]_k}{[j]_k [n+1-i]_k}.$$

This is a strictly decreasing sequence in k , and Theorem 2.1 yields

$$\text{Corr}(X_{i:n}, X_{j:n}) \leq \sqrt{\rho_1^2} = \sqrt{\frac{i(n+1-j)}{j(n+1-i)}},$$

with equality if and only if $g(u) (= F^{-1}(u)) = \alpha u + \beta$ for some $\alpha > 0$ and $\beta \in \mathbb{R}$, i.e., $X \sim \mathcal{U}(\beta, \beta + \alpha)$, $\alpha > 0$.

The same simple arguments apply to the case where (X, Y) has a density as in (2). Then, it is easily seen that for any fixed x and y in $(0, 1)$,

$$(X|Y = y) \stackrel{d}{=} yB_{\alpha, \beta} \quad \text{and} \quad (Y|X = x) \stackrel{d}{=} x + (1-x)B_{\beta, \gamma} \stackrel{d}{=} 1 - B_{\gamma, \beta} + xB_{\gamma, \beta},$$

where $B_{r,s}$ denotes a Beta random variable with parameters $r > 0$ and $s > 0$. It follows that

$$\mathbb{E}(X^n|Y) = A_n Y^n \quad \text{and} \quad \mathbb{E}(Y^n|X) = B_n X^n + \text{Pol}_{n-1}(X)$$

with

$$A_n = \mathbb{E}[B_{\alpha, \beta}^n] = \frac{[\alpha]_n}{[\alpha + \beta]_n} \quad \text{and} \quad B_n = \mathbb{E}[B_{\gamma, \beta}^n] = \frac{[\gamma]_n}{[\beta + \gamma]_n}.$$

Since $\rho_n^2 = A_n B_n = \frac{[\alpha]_n [\gamma]_n}{[\alpha+\beta]_n [\beta+\gamma]_n}$ is a strictly decreasing function in n , Theorem 2.1 yields $R(X, Y) = |\rho_1| = \rho_1 = \rho(X, Y)$, which is (3).

Nevzorov's characterization of exponential distribution. Nevzorov (1992) proved that for any $n, m \in \{1, 2, \dots\}$,

$$\text{Corr}(R_n, R_{n+m}) \leq \sqrt{\frac{n}{n+m}},$$

where R_i is the i -th (upper) record from a continuous distribution F with finite variance (here $R_1 = X_1$ is the first observed random variable in the iid sequence $\{X_i\}_{i=1}^\infty$). Moreover, equality characterizes the location-scale family of the standard exponential distribution. Theorem 2.1 gives the result immediately. Indeed, if W_i denotes the i -th record from $\mathcal{Exp}(1)$ (with density $f(x) = e^{-x}$, $x > 0$) then

$$(W_n, W_{n+m}) \stackrel{d}{=} (E_1 + \dots + E_n, E_1 + \dots + E_{n+m}), \quad n, m \in \{1, 2, \dots\},$$

where $\{E_i\}_{i=1}^\infty$ is an iid sequence from $\mathcal{Exp}(1)$ – see, e.g., Arnold et al. (1998). Setting $X = E_1 + \dots + E_n$ and $Y = E_1 + \dots + E_{n+m}$, the joint density of (X, Y) is

$$f_{X,Y}(x, y) = \frac{1}{\Gamma(n)\Gamma(m)} x^{n-1} (y-x)^{m-1} e^{-y}, \quad 0 < x < y < \infty,$$

and the conditional densities are

$$f_{X|Y}(x|y) = \frac{\Gamma(n+m)}{\Gamma(n)\Gamma(m)} x^{n-1} (y-x)^{m-1} y^{-(n+m-1)}, \quad x \in (0, y),$$

and

$$f_{Y|X}(y|x) = \frac{1}{\Gamma(m)} (y-x)^{m-1} e^{-(y-x)}, \quad y \in (x, \infty).$$

It follows that

$$\mathbb{E}(X^k | Y = y) = \frac{(k+n-1)!(n+m-1)!}{(k+n+m-1)!(n-1)!} y^k$$

and

$$\mathbb{E}(Y^k | X = x) = x^k + \frac{1}{\Gamma(m)} \sum_{i=1}^k \binom{k}{i} \Gamma(i+m) x^{k-i}.$$

Thus, A3 is satisfied with $A_k = \frac{(k+n-1)!(n+m-1)!}{(k+n+m-1)!(n-1)!}$ and $B_k = 1$, so that

$$\rho_k^2 = A_k B_k = \frac{(k+n-1)!(n+m-1)!}{(k+n+m-1)!(n-1)!} = \frac{[n]_k}{[n+m]_k}.$$

Since this is a strictly decreasing sequence in k , Theorem 2.1 yields the inequality

$$\text{Corr}(R_n, R_{n+m}) = \text{Corr}(g(W_n), g(W_{n+m})) \leq \sqrt{\rho_1^2} = \sqrt{\frac{n}{n+m}},$$

where $g(u) = F^{-1}(1 - e^{-u})$, $u > 0$. The equality holds if and only if g is increasing and linear, that is, if and only if F is the distribution of $\alpha E + \beta$ where $\alpha > 0$, $\beta \in \mathbb{R}$ and $E \sim \mathcal{Exp}(1)$.

López-Blázquez and Castaño-Martínez result on maximal correlation of order statistics from a finite population. Let $U_{1:n}^{(N)} < U_{2:n}^{(N)} < \dots < U_{n:n}^{(N)}$ be the order statistics corresponding to a simple random sample, $U_1^{(N)}, \dots, U_n^{(N)}$, taken without replacement from the finite ordered population $\Pi_N = \{1, 2, \dots, N\}$, where $2 \leq n < N$. Since $\mathbb{P}(U_{i:n}^{(N)} = k) = \binom{k-1}{i-1} \binom{N-k}{n-i} \binom{N}{n}^{-1}$ for $k \in \{i, i+1, \dots, N - (n-i)\}$ (and 0 otherwise), and this defines a probability mass function with support $A_{i:n}^{(N)} := \{i, i+1, \dots, N - (n-i)\}$, we conclude the identity

$$\sum_{k=i}^{N-(n-i)} \binom{k-1}{i-1} \binom{N-k}{n-i} = \binom{N}{n}, \quad 1 \leq i \leq n \leq N. \quad (14)$$

Setting $[\alpha]_m = \alpha(\alpha+1) \cdots (\alpha+m-1)$ (with $[\alpha]_0 = 1$ for all $\alpha \in \mathbb{R}$), we can derive, with the help of (14), a simple expression for the ascending moments of $U_{i:n}^{(N)}$:

$$\mathbb{E} \left\{ [U_{i:n}^{(N)}]_m \right\} = [N+1]_m \frac{[i]_m}{[n+1]_m}, \quad m = 1, 2, \dots \quad (15)$$

We also mention the following obvious relations, holding for all $1 \leq i < j \leq n$:

$$(U_{i:n}^{(N)}, U_{j:n}^{(N)}) \stackrel{d}{=} (N+1 - U_{n+1-i:n}^{(N)}, N+1 - U_{n+1-j:n}^{(N)}), \quad (16)$$

$$(U_{i:n}^{(N)} | U_{j:n}^{(N)} = s) \stackrel{d}{=} U_{i:j-1}^{(s-1)}, \quad s \in \{j, j+1, \dots, N - (n-j)\}, \quad (17)$$

$$(U_{j:n}^{(N)} | U_{i:n}^{(N)} = k) \stackrel{d}{=} k + U_{j-i:n-i}^{(N-k)}, \quad k \in \{i, i+1, \dots, N - (n-i)\}. \quad (18)$$

Now, by (15) and (17) we get

$$\mathbb{E} \left\{ [U_{i:n}^{(N)}]_m \middle| U_{j:n}^{(N)} = s \right\} = [s]_m \frac{[i]_m}{[j]_m}, \quad m = 1, 2, \dots \quad (19)$$

Let $(X, Y) = (U_{i:n}^{(N)}, U_{j:n}^{(N)})$. Relation (19) shows that

$$\mathbb{E}([X]_m | Y) = \frac{[i]_m}{[j]_m} [Y]_m = \frac{[i]_m}{[j]_m} Y^m + \text{Pol}_{m-1}(Y), \quad m = 1, 2, \dots,$$

and this implies, using induction on m , that

$$\mathbb{E}(X^m | Y) = \frac{[i]_m}{[j]_m} Y^m + \text{Pol}_{m-1}(Y), \quad m = 1, 2, \dots \quad (20)$$

Similarly, setting $i' = n + 1 - j$, $j' = n + 1 - i$ (so that $1 \leq i' < j' \leq n$), writing $U_{i'}$ instead of $U_{i':n}^{(N)}$, $U_{j'}$ instead of $U_{j':n}^{(N)}$, and applying relations (16) and (19), we get

$$\begin{aligned}
\mathbb{E}([Y]_m | X = k) &= \mathbb{E}\{[N + 1 - U_{i'}]_m | U_{j'} = N + 1 - k\} \\
&= \mathbb{E}\{(-1)^m [U_{i'}]_m + \text{Pol}_{m-1}(U_{i'}) | U_{j'} = N + 1 - k\} \\
&= (-1)^m \mathbb{E}\{[U_{i'}]_m | U_{j'} = N + 1 - k\} + \text{Pol}_{m-1}(N + 1 - k) \\
&= (-1)^m [N + 1 - k]_m \frac{[i']_m}{[j']_m} + \text{Pol}_{m-1}(k) \\
&= [k]_m \frac{[i']_m}{[j']_m} + \text{Pol}_{m-1}(k) = [k]_m \frac{[n + 1 - j]_m}{[n + 1 - i]_m} + \text{Pol}_{m-1}(k).
\end{aligned}$$

It follows that $\mathbb{E}([Y]_m | X) = \frac{[n+1-j]_m}{[n+1-i]_m} [X]_m + \text{Pol}_{m-1}(X) = \frac{[n+1-j]_m}{[n+1-i]_m} X^m + \text{Pol}_{m-1}(X)$ and, finally, using induction on m , we get the expression

$$\mathbb{E}(Y^m | X) = \frac{[n + 1 - j]_m}{[n + 1 - i]_m} X^m + \text{Pol}_{m-1}(X), \quad m = 1, 2, \dots \quad (21)$$

Clearly, (20) and (21) show that A3 is satisfied for (X, Y) . Moreover, we have found that $A_m = \frac{[i]_m}{[j]_m}$ and $B_m = \frac{[n+1-j]_m}{[n+1-i]_m}$ (both do not dependent on N). Hence, since $\rho_m^2 = A_m B_m$ is a strictly decreasing sequence in m , we obtain from Theorem 2.1 the inequality

$$\text{Corr}[g_1(U_{i:n}^{(N)}), g_2(U_{j:n}^{(N)})] \leq \sqrt{\rho_1^2} = \sqrt{\frac{i(n+1-j)}{j(n+1-i)}},$$

in which the equality holds if and only if both g_1 and g_2 are (non-constant and) linear and with the same monotonicity – more precisely, the restriction of g_1 in the set $A_{i:n}^{(N)}$ has to be non-constant and linear and the restriction of g_2 in the set $A_{j:n}^{(N)}$ has to be non-constant and linear and with the same monotonicity as g_1 ; note that both sets $A_{i:n}^{(N)}$ and $A_{j:n}^{(N)}$ contain at least two points if and only if $N \geq n + 1$. Lemma 2.1 of Balakrishnan et al. (2003) asserts that for the non-decreasing function $g : \{1, 2, \dots, N\} \rightarrow \{x_1 \leq x_2 \leq \dots \leq x_N\} := \tilde{\Pi}_N$ with $g(i) = x_i$, $i = 1, 2, \dots, N$, it is true that

$$(g(U_{i:n}^{(N)}), g(U_{j:n}^{(N)})) \stackrel{d}{=} (X_{i:n}, X_{j:n}), \quad 1 \leq i < j \leq n,$$

where $X_{1:n} \leq X_{2:n} \leq \dots \leq X_{n:n}$ are the order statistics corresponding to a simple random sample drawn (without replacement) from the finite population $\tilde{\Pi}_N$. Suppose that $\text{Corr}(X_{i:n}, X_{j:n})$ is well-defined or, equivalently, that the elements of $\tilde{\Pi}_N$ satisfy $x_i < x_{N-(n-i)}$ and $x_j < x_{N-(n-j)}$ (otherwise, at least one of $X_{i:n}, X_{j:n}$ would be degenerate). Then we conclude that

$$\text{Corr}(X_{i:n}, X_{j:n}) \leq \sqrt{\frac{i(n+1-j)}{j(n+1-i)}}, \quad 1 \leq i < j \leq n < N, \quad (22)$$

and the equality (for fixed i, j, n, N) characterizes those finite populations $\tilde{\Pi}_N$ for which the sets $\{x_i, x_{i+1}, \dots, x_{N-(n-i)}\}$ and $\{x_j, x_{j+1}, \dots, x_{N-(n-j)}\}$ (that may or

may not have common points) consist of consecutive terms of two (possibly different) strictly increasing arithmetic progresses. That is, a population of size N with elements $x_1 \leq x_2 \leq \dots \leq x_N$ satisfying $x_i < x_{N-(n-i)}$ and $x_j < x_{N-(n-j)}$ attains the equality in (22) if and only there exist constants $a_1 > 0$, $b_1 \in \mathbb{R}$, $a_2 > 0$ and $b_2 \in \mathbb{R}$ such that

$$x_k = \begin{cases} a_1 k + b_1, & \text{for } k = i, i+1, \dots, N-(n-i), \\ a_2 k + b_2, & \text{for } k = j, j+1, \dots, N-(n-j), \\ \text{arbitrary,} & \text{otherwise.} \end{cases}$$

López-Blázquez and Castaño-Martínez (2006), using Hahn polynomials, have obtained a corresponding inequality for the correlation ratio, which implies inequality (22); their arguments, however, apply to populations $\tilde{\Pi}_N$ having N distinct elements. We also refer to Theorem 2.1 and Corollary 2.1 in Castaño-Martínez et al. (2007), noting that the characterization result stated in Corollary 2.1 of this article is incomplete, unless the sets $A_{i:n}^{(N)}$ and $A_{j:n}^{(N)}$ have at least two common points, i.e., $N \geq n + (j - i) + 1$.

4 Records from a splitting model and a Nevzorov-type characterization of the exponential distribution

Assume that in a particular country and for a specific athletic event, the consecutive performances of the athletes are described by an iid sequence $\{X_i\}_{i=1}^\infty$. Here and elsewhere in this section, the common distribution of each X_i will be assumed to be continuous, i.e., with no atoms – absolute continuity is not needed. As the time goes on, the common practice is that some data regarding the sequence of national records, i.e., the sequence $\{R_i\}_{i=1}^\infty$, are saved (and recorded), in contrast to the original performances of the athletes, X_i , which are usually lost or forgotten. The above considerations give rise to the classical record model (based on an iid sequence), which is well-developed in the literature; see, e.g., Arnold et al. (1998). Under this classical model the observed sequence $\{R_i\}_{i=1}^n$ of the first n upper national records is defined as $R_1 = X_1$ and $R_i = X_{T(i)}$, $i = 2, \dots, n$, where $T(i) = \min\{j \in \{1, 2, \dots\} : X_j > R_{i-1}\}$.

Suppose now that, after the appearance of the n -th national record, the initial country is divided into (say) two new countries (branches), and assume that the athletes in each country are of the same strength as they was before the division. Then, the subsequent national records in each branch will take under account the current (common) national record, R_n , and the subsequent sequence of their individual records will be of the form $(R'_{n+n_1}, R''_{n+n_2})$, with $n_1, n_2 \in \{1, 2, \dots\}$. Clearly,

$$R'_{n+n_1} \stackrel{d}{=} R_{n+n_1} \quad \text{and} \quad R''_{n+n_2} \stackrel{d}{=} R_{n+n_2} \quad (23)$$

where R_{n+m} is the $(n+m)$ -th record from the initial sequence, but as n_1 and n_2 become large, the random variables R'_{n+n_1} and R''_{n+n_2} should tend to be independent.

Thus, the actual definition of the splitting record sequence is equivalent to the following model: Let $\{X_1, X'_1, X''_1, X_2, X'_2, X''_2, \dots\}$ be an iid sequence of random variables. Define the n -th upper record R_n as before (based on the X_i 's), then set $R'_n = R''_n := R_n$ and $T'(n) = T''(n) := T(n)$, and for $i = 1, 2, \dots$ define the subsequent record times and record values by

$$\begin{aligned} T'(n+i) &= \min\{j \in \{1, 2, \dots\} : X'_j > R'_{n+i-1}\}, \quad R'_{n+i} = X'_{T'(n+i)}, \quad \text{and} \\ T''(n+i) &= \min\{j \in \{1, 2, \dots\} : X''_j > R''_{n+i-1}\}, \quad R''_{n+i} = X''_{T''(n+i)}. \end{aligned}$$

Clearly, it is of some interest to study the correlation behavior of the marginal records under this model, since large correlation among these variables entails good prediction of the one branch to the other. It is not surprising that, similarly to the classic case, the splitting record sequence satisfies several interesting properties. In particular, in the sequel we shall make use of the following lemma, the proof of which is simple and is left to the reader – cf. Arnold et al. (1998).

Lemma 4.1. (a) If $\{(W'_{n+n_1}, W''_{n+n_2})\}_{n_1, n_2=1}^\infty$ is the splitting record sequence based on the iid sequence $\{E_i, E'_i, E''_i\}_{i=1}^\infty$ from the standard exponential distribution, $\mathcal{E}\text{xp}(1)$, then for each $n_1, n_2 \in \{1, 2, \dots\}$,

$$(W'_{n+n_1}, W''_{n+n_2}) \stackrel{d}{=} (E_1 + \dots + E_n + E'_1 + \dots + E'_{n_1}, E_1 + \dots + E_n + E''_1 + \dots + E''_{n_2}). \quad (24)$$

(b) Let $\{(R'_{n+n_1}, R''_{n+n_2})\}_{n_1, n_2=1}^\infty$ be the splitting record sequence based on an iid sequence $\{X_i, X'_i, X''_i\}_{i=1}^\infty$ from a non-atomic (continuous) distribution function F . Then, for each $n_1, n_2 \in \{1, 2, \dots\}$,

$$(R'_{n+n_1}, R''_{n+n_2}) \stackrel{d}{=} (g(W'_{n+n_1}), g(W''_{n+n_2})), \quad (25)$$

where $g(u) = F^{-1}(1 - e^{-u})$, $u > 0$, with $F^{-1}(y) = \inf\{x : F(x) \geq y\}$, $y \in (0, 1)$.

With the help of Lemma 4.1, Theorem 2.1 yields the following characterization.

Theorem 4.1. If $(R'_{n+n_1}, R''_{n+n_2})$ are splitting records based on an iid sequence $\{X_i\}_{i=1}^\infty$ from a non-atomic distribution F with $\mathbb{E}(R'_{n+n_1})^2 < \infty$ and $\mathbb{E}(R''_{n+n_2})^2 < \infty$ then

$$\text{Corr}(R'_{n+n_1}, R''_{n+n_2}) \leq \frac{n}{\sqrt{n+n_1}\sqrt{n+n_2}},$$

and the equality holds if and only if F is the distribution function of $\alpha E + \beta$ for some $\alpha > 0$ and $\beta \in \mathbb{R}$, where $E \sim \mathcal{E}\text{xp}(1)$.

Proof: Set $X = E_1 + \dots + E_n + E'_1 + \dots + E'_{n_1}$ and $Y = E_1 + \dots + E_n + E''_1 + \dots + E''_{n_2}$ with (E_1, \dots, E''_{n_2}) being a vector of $n + n_1 + n_2$ iid standard exponential rv's. It can be seen (see the proof of Theorem 4.2, below) that for all $k \in \{1, 2, \dots\}$,

$$\mathbb{E}(X^k|Y) = \frac{[n]_k}{[n+n_2]_k} Y^k + \text{Pol}_{k-1}(Y), \quad \mathbb{E}(Y^k|X) = \frac{[n]_k}{[n+n_1]_k} X^k + \text{Pol}_{k-1}(X).$$

That is, the random vector (X, Y) has the polynomial regression property with $A_k = [n]_k/[n+n_2]_k$ and $B_k = [n]_k/[n+n_1]_k$. Clearly, $\rho_k^2 = ([n]_k)^2/([n+n_1]_k[n+n_2]_k)$ is strictly decreasing in k . In view of Lemma 4.1, Theorem 2.1 shows that, with $g(u) = F^{-1}(1 - e^{-u})$,

$$\begin{aligned}\text{Corr}(R'_{n+n_1}, R''_{n+n_2}) &= \text{Corr}(g(W'_{n+n_1}), g(W''_{n+n_2})) \\ &= \text{Corr}(g(X), g(Y)) \leq \sqrt{\rho_1^2} = \frac{n}{\sqrt{n+n_1}\sqrt{n+n_2}},\end{aligned}$$

and the equality holds if and only if $g : (0, \infty) \rightarrow \mathbb{R}$ is linear; this, together with the fact that g has assumed to be strictly increasing, completes the proof. \square

Theorem 4.1 and Nevzorov's (1992) characterization reflects the polynomial regression property of a specific class of multivariate Gamma random vectors, provided that every component is representative as a sum on independent Gamma rv's with the same scale parameter, say $1/\lambda$. Recall that a random variable X follows a Gamma distribution with parameters $\alpha > 0$ and $\lambda > 0$ if its density is given by

$$f(x) = \frac{\lambda^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\lambda x}, \quad x > 0;$$

this fact is denoted by $X \sim \Gamma(\alpha; \lambda)$, while the notation $X \sim \Gamma(0; \lambda)$ (for some $\lambda > 0$) means that X is degenerate and takes the value zero w.p. 1. In any case, $\mathbb{E}X = \alpha/\lambda$ and $\text{Var} X = \alpha/\lambda^2$. Under the above notation one can easily verify the following result, which essentially contains both Theorem 4.1 and Nevzorov's characterization as particular cases.

Theorem 4.2. Let $X_i \sim \Gamma(\alpha_i; \lambda)$ ($i = 0, 1, 2$) be independent rv's with $\lambda > 0$, $\alpha_i \geq 0$ ($i = 0, 1, 2$) and $\alpha_0 + \alpha_i > 0$ ($i = 1, 2$). Then the random vector $(X, Y) = (X_0 + X_1, X_0 + X_2)$ follows a bivariate distribution with Gamma marginals, namely $X \sim \Gamma(\alpha_0 + \alpha_1; \lambda)$ and $Y \sim \Gamma(\alpha_0 + \alpha_2; \lambda)$. Moreover, (X, Y) has the polynomial regression property: For all $n \in \{1, 2, \dots\}$,

$$\mathbb{E}(X^n|Y) = \sum_{j=0}^n \binom{n}{j} \frac{[\alpha_0]_j [\alpha_1]_{n-j}}{\lambda^{n-j} [\alpha_0 + \alpha_2]_j} Y^j, \quad \mathbb{E}(Y^n|X) = \sum_{j=0}^n \binom{n}{j} \frac{[\alpha_0]_j [\alpha_2]_{n-j}}{\lambda^{n-j} [\alpha_0 + \alpha_1]_j} X^j,$$

where $[\alpha]_0 \equiv 1$ for all $\alpha \in \mathbb{R}$ and $[\alpha]_k = \alpha(\alpha+1) \cdots (\alpha+k-1)$ ($k = 1, 2, \dots$). Finally, for any $g_1 \in L^2(X)$ with $\text{Var} g_1(X) > 0$ and for any $g_2 \in L^2(Y)$ with $\text{Var} g_2(Y) > 0$ we have the inequality

$$\text{Corr}(g_1(X), g_2(Y)) \leq \frac{\alpha_0}{\sqrt{\alpha_0 + \alpha_1} \sqrt{\alpha_0 + \alpha_2}},$$

where, provided that $\alpha_1 + \alpha_2 > 0$, the equality holds if and only if either $\alpha_0 = 0$ (and g_1, g_2 are arbitrary) or $\alpha_0 > 0$ and both g_1, g_2 are nonconstant, linear and with the same monotonicity.

Proof: Cases $\alpha_0 = 0$ and $\alpha_1 = \alpha_2 = 0$ are trivial (X, Y are independent and $X = Y$ w.p. 1, respectively). Both cases $\alpha_0 > 0, \alpha_1 = 0, \alpha_2 > 0$ and $\alpha_0 > 0, \alpha_1 > 0, \alpha_2 = 0$ are similar to Nevzorov's case and can be shown as in Section 3. Assume now that $\alpha_i > 0$ for $i = 0, 1, 2$. Then, it is easily seen that the conditional density of X given $Y = y$ (for any fixed $y > 0$) is given by

$$f_{X|Y}(x|y) = ce^{-\lambda x} \int_0^{\min\{x,y\}} w^{\alpha_0-1} (x-w)^{\alpha_1-1} (y-w)^{\alpha_2-1} e^{\lambda w} dw, \quad x > 0,$$

where

$$c = c(\alpha_0, \alpha_1, \alpha_2; \lambda; y) = \frac{\lambda^{\alpha_1} \Gamma(\alpha_0 + \alpha_2)}{y^{\alpha_0 + \alpha_2 - 1} \Gamma(\alpha_0) \Gamma(\alpha_1) \Gamma(\alpha_2)}.$$

Despite the fact that this conditional density is not given in a closed form, we can calculate $\mathbb{E}(X^n|Y = y)$ using Tonelli's Theorem. Indeed, consider the nonnegative functions $\theta(w) = w^{\alpha_0-1} e^{\lambda w}$ ($w > 0$) and $h(x, y, w) = (x-w)^{\alpha_1-1} (y-w)^{\alpha_2-1} I(w < \min\{x, y\})$ ($x, y, w > 0$). Then,

$$\begin{aligned} \mathbb{E}(X^n|Y = y) &= c \left\{ \int_0^y x^n e^{-\lambda x} \int_0^x \theta(w) h(x, y, w) dw dx \right. \\ &\quad \left. + \int_y^\infty x^n e^{-\lambda x} \int_0^y \theta(w) h(x, y, w) dw dx \right\} \\ &= c \left\{ \int_0^y \theta(w) \int_w^y x^n e^{-\lambda x} h(x, y, w) dx dw \right. \\ &\quad \left. + \int_0^y \theta(w) \int_y^\infty x^n e^{-\lambda x} h(x, y, w) dx dw \right\} \\ &= c \int_0^y \theta(w) \int_w^\infty x^n e^{-\lambda x} h(x, y, w) dx dw \\ &= c \int_0^y w^{\alpha_0-1} (y-w)^{\alpha_2-1} \left\{ \int_0^\infty (x+w)^n e^{-\lambda x} x^{\alpha_1-1} dx \right\} dw. \end{aligned}$$

Now, expanding $(x+w)^n$ according to Newton's formula and using $\int_0^\infty x^{j+\alpha_1-1} e^{-\lambda x} dx = \Gamma(\alpha_1 + j)/\lambda^{\alpha_1+j}$ ($j = 0, 1, \dots, n$) we get for the inner integral the expression

$$\int_0^\infty (x+w)^n e^{-\lambda x} x^{\alpha_1-1} dx = \frac{1}{\lambda^{\alpha_1}} \sum_{j=0}^n \binom{n}{j} \frac{\Gamma(\alpha_1 + j)}{\lambda^j} w^{n-j}.$$

Finally, substituting this expression to the double integral, above, we get

$$\begin{aligned} \mathbb{E}(X^n|Y = y) &= \frac{c}{\lambda^{\alpha_1}} \sum_{j=0}^n \binom{n}{j} \frac{\Gamma(\alpha_1 + j)}{\lambda^j} \int_0^y w^{\alpha_0 + (n-j)-1} (y-w)^{\alpha_2-1} dw \\ &= \frac{c}{\lambda^{\alpha_1}} \sum_{j=0}^n \binom{n}{j} \frac{\Gamma(\alpha_1 + j)}{\lambda^j} \frac{\Gamma(\alpha_2) \Gamma(\alpha_0 + n - j)}{\Gamma(\alpha_0 + \alpha_2 + n - j)} y^{\alpha_0 + \alpha_2 + (n-j)-1} \\ &= \frac{\Gamma(\alpha_0 + \alpha_2)}{\Gamma(\alpha_0) \Gamma(\alpha_1)} \sum_{j=0}^n \binom{n}{j} \frac{\Gamma(\alpha_0 + j) \Gamma(\alpha_1 + n - j)}{\lambda^{n-j} \Gamma(\alpha_0 + \alpha_2 + j)} y^j. \end{aligned}$$

Therefore, X has polynomial regression on Y and, similarly, Y has polynomial regression on X . It follows that (X, Y) satisfies conditions A1–A3 and, moreover,

$$\rho_n = \text{sign}(A_n) \sqrt{A_n B_n} = \frac{[\alpha_0]_n}{\sqrt{[\alpha_0 + \alpha_1]_n} \sqrt{[\alpha_0 + \alpha_2]_n}};$$

since $|\rho_n| = \rho_n$ is strictly decreasing in n , an application of Theorem 2.1 completes the proof. \square

Theorem 4.2 includes Nevzorov’s (1992) characterization because, taking $\lambda = 1$, $\alpha_0 = n$, $\alpha_1 = 0$, $\alpha_2 = m$ and $g_1(u) = g_2(u) = F^{-1}(1 - e^{-u})$, $u > 0$, we have that, under the standard record model, $(R_n, R_{n+m}) \stackrel{d}{=} (g(W_n), g(W_{n+m})) \stackrel{d}{=} (g(X), g(Y))$, where (W_n, W_{n+m}) are the corresponding upper records from the standard exponential distribution. Clearly, it also includes the result on splitting record models of Theorem 4.1 – the only difference being that, due to Lemma 4.1, one has now to put $\alpha_1 = n_1$ (rather than $\alpha_1 = 0$) and $\alpha_2 = n_2$ (rather than $\alpha_2 = m$).

Also, it is of some interest to note that (11) yields the covariance identity (cf. Afendras et al. (2011))

$$\text{Cov}[g_1(X), g_2(Y)] = \sum_{n=1}^{\infty} \frac{[\alpha_0]_n}{n! [\alpha_0 + \alpha_1]_n [\alpha_0 + \alpha_2]_n} \mathbb{E}[X^n g_1^{(n)}(X)] \mathbb{E}[Y^n g_2^{(n)}(Y)], \quad (26)$$

provided that $g_1, g_2 \in C^\infty(0, \infty)$, $g_1(X) \in L^2(X)$, $g_2(Y) \in L^2(Y)$, and assuming that $\mathbb{E}|X^n g_1^{(n)}(X)| < \infty$ and $\mathbb{E}|Y^n g_2^{(n)}(Y)| < \infty$ for all n , where $g_i^{(n)}$ denotes the n -th derivative of g_i , $i = 1, 2$. Of course one can apply (26) to the case $\alpha_1 = \alpha_2 = 0$, $\alpha_0 > 0$; then, $X = Y \sim \Gamma(\alpha_0; \lambda)$ and we get the (known) generalized Stein-type identity for the $\Gamma(\alpha_0; \lambda)$ distribution:

$$\text{Cov}[g_1(X), g_2(X)] = \sum_{n=1}^{\infty} \frac{1}{n! [\alpha_0]_n} \mathbb{E}[X^n g_1^{(n)}(X)] \mathbb{E}[X^n g_2^{(n)}(X)]. \quad (27)$$

Also, we can apply (26) to the classical record setup from the standard exponential (on taking $\lambda = 1$, $\alpha_0 = n$, $\alpha_1 = 0$ and $\alpha_2 = m$); then we get the identity

$$\text{Cov}[g_1(W_n), g_2(W_{n+m})] = \sum_{k=1}^{\infty} \frac{1}{k! [n+m]_k} \mathbb{E}[W_n^k g_1^{(k)}(W_n)] \mathbb{E}[W_{n+m}^k g_2^{(k)}(W_{n+m})]. \quad (28)$$

5 Conclusions

It is clear that the simplicity of the proposed method depends heavily on the polynomial regression property, A3, which is satisfied by all bivariate distributions discussed in the present article. Castaño-Martínez et al. (2007) developed a correlation model for partial minima (or maxima) rather than records. Their Section 3 indicates that many difficulties can enter to the correlation problem when A3 fails; it seems that, in such cases, one has to calculate the values of $\rho_{n,k} = \mathbb{E}[\phi_n(X)\psi_k(Y)]$ for all n and k . This is not an easy task in general, in contrast to the present simplified case, where knowledge of the values A_n and B_n in A3 suffices for calculating the maximal correlation coefficient.

REFERENCES

- Afendras, G.; Papadatos, N.; Papathanasiou, V. (2011). An extended Stein-type covariance identity for the Pearson family, with applications to lower variance bounds, *Bernoulli*, **17**, 507–529.
- Arnold, B.C.; Balakrishnan, N.; Nagaraja, H.N. (1998). *Records*. Wiley, New York.
- Balakrishnan, N.; Charalambides, C.; Papadatos, N. (2003). Bounds on expectation of order statistics from a finite population, *J. Statist. Plann. Inference*, **113**, 569–588.
- Breiman, L.; Friedman, J.H. (1985). Estimating optimal transformations for multiple regression and correlation, *J. Amer. Statist. Assoc.*, **80**, 590–598.
- Bryc, W.; Dembo, A.; Kagan, A. (2005). On the maximum correlation coefficient, *Theory Probab. Appl.*, **49**, 132–138.
- Castañó-Martínez, A.; López-Blázquez, F.; Salamanca-Miño, B. (2007). Maximal correlation between order statistics. In: *Recent Developments in Ordered Random Variables*, M. Ahsanullah and M. Raqab (eds.), Nova Science Publishers, 55–68.
- Dembo, A.; Kagan, A.; Shepp, L.A. (2001). Remarks on the maximum correlation coefficient, *Bernoulli*, **7**, 343–350.
- Gebelein, H. (1941). Das Statistische Problem der Korrelation als Variation und Eigenwertproblem und sein Zusammenhang mit der Ausgleichrechnung, *Angew. Math. Mech.*, **21**, 364–379.
- Koudou, A.E. (1996). Probabilités de Lancaster. *Exp. Math.*, **14**, 247–275.
- Koudou, A.E. (1998). Lancaster bivariate probability distributions with Poisson, negative binomial and gamma margins. *Test*, **7**, 95–110.
- Lancaster, H.O. (1957). Some properties of the bivariate normal distribution considered in the form of a contingency table, *Biometrika*, **44**, 289–292.
- Lancaster, H.O. (1969). *The chi-squared distribution*, Wiley, New York.
- Liu, J.S.; Wong, W.H.; Kong, A. (1994). Covariance structure of the Gibbs sampler with applications to the comparisons of estimators and augmentation schemes, *Biometrika*, **81**, 27–40.
- López-Blázquez, F.; Castañó-Martínez, A. (2006). Upper and lower bounds for the correlation ratio of order statistics from a sample without replacement, *J. Statist. Plann. Inference*, **136**, 43–52.
- Nevzorov, V.B. (1992). A characterization of exponential distributions by correlations between records, *Math. Meth. Statist.*, **1**, 49–54.
- Székely, G.J.; Móri, T.F. (1985). An extremal property of rectangular distributions, *Statist. Probab. Lett.*, **3**, 107–109.
- Terrell, G.R. (1983). A characterization of rectangular distributions, *Ann. Probab.*, **11**, 823–826.
- Yu, Y. (2008). On the maximal correlation coefficient, *Statist. Probab. Lett.*, **78**, 1072–1075.